Journal of Informatics Management and Information Technology
ISSN 2774-4744 (Media Online)
Vol 5, No 2, April 2025| Hal 95-103
DOI: 10.47065/jimat.v5i2.478
https://hostjournals.com/jimat

Pemilihan Fitur Menggunakan Chi-Square Untuk Deteksi Serangan Pada Jaringan Internet of Medical Things Menggunakan Random Forest

Alfia Tiara Permatasari[1], Rusdiyanto[2], M Agus Syamsul Arifin[1,*]

[1] Fakultas Ilmu Teknik, Program Studi Rekayasa Sistem Komputer, Universitas Bina Insan, Kota Lubuklinggau , Indonesia
[2] Fakultas Ilmu Teknik, Program Studi Informatika, Universitas Bina Insan, Kota Lubuklinggau , Indonesia
Jl. Jendral H.M Soeharto, Kelurahan Lubuk Kupang, Kecamatan Lubuklinggau Selatan I, Kota Lubukinggau, Indonesia
Email: [1] 2102010005@mhs.univbinainsan.ac.id, [2] rusdiyanto@univbinainsan.ac.id, [3,*]mas.agus1988@gmail.com
Email Penulis Korespondensi: mas.agus1988@gmail.com

Abstrak−Penelitian ini membahas tentang keamanan jaringan Internet of Medical Things (IoMT) dengan menggunakan dataset CICIoMT2024 untuk menganalisis dan mendeteksi serangan siber melalui penerapan metode Machine Learning (ML). Penelitian ini menggunakan teknik pemilihan fitur Chi-Square untuk mengidentifikasi fitur-fitur penting, serta menggunakan algoritma Random Forest untuk proses klasifikasi data. Pemanfaatan fitur Chi-Square khususnya dalam analisis lalu lintas jaringan perangkat IoMT belum banyak dieksplorasi, sehingga penelitian ini memberikan kontribusi baru dalam bidang tersebut. Hasil dari pemilihan fitur Chi-Square digunakan untuk melatih model Machine Learning dalam mengklasifikasikan data antara lalu lintas normal dan lalu lintas yang mengandung serangan. Dalam eksperimen yang dilakukan, algoritma Random Forest menunjukkan performa yang sangat baik dengan mencapai akurasi hingga 100%, serta nilai precision, recall, dan F1-Score yang tinggi. Hasil ini menunjukkan bahwa Random Forest mampu menangani kompleksitas data IoMT secara efektif. Dengan demikian, dapat disimpulkan bahwa algoritma Random Forest sangat relevan dan efektif digunakan dalam penelitian keamanan jaringan IoMT.

Kata Kunci: IoMT; CICIoMT2024; Chi-Square.

Abstract−This research discusses the security of the Internet of Medical Things (IoMT) network using the CICIoMT2024 dataset to analyze and detect cyber attacks through the application of Machine Learning (ML) methods. This research uses the Chi-Square feature selection technique to identify important features, and uses the Random Forest algorithm for the data classification process. The utilization of Chi-Square features, especially in the analysis of network traffic of IoT devices, has not been widely explored, so this research makes a new contribution to the field. The result of the Chi-Square feature selection are used to train Machine Learning models to classify data between normal traffic and traffic containing attacks. In the experiments conducted, the Random Forest algorithm showed excellent performance by achieving up to 100% accuracy, as well as high precision, recall, F1-Score values. These results show that Random Forest is able to handle the complexity of IoMT data effectively. Thus, it can be concluded that the Random Forest algorithm is very relevant and effective to use in IoMT network security research.

Keywords: IoMT; CICIoMT2024; Chi-Square

1. PENDAHULUAN

Internet of Medical Things (IoMT) merupakan jaringan perangkat medis yang saling terhubung melalui internet dan merupakan evolusi dari Internet of things (IoT). Istilah internet of medical things pertama kali diperkenalkan oleh Kevin Ashton pada tahun 1999 di Massachusetts Institute of Technology (MIT), meskipun penerapannya dalam bidang medis baru berkembang pada awal tahun 2000-an, seiring dengan kemajuan teknologi sensor dan komunikasi nirkabel[1]. Perkembangan ini memungkinkan IoT digunakan dalam berbagai sektor, termasuk kesehatan. Salah satu penerapan awal IoMT adalah pemantauan kondisi pasien secara real-time melalui sensor yang dapat mengirimkan data langsung kepada tenaga medis melalui jaringan internet[2]. Dalam sistem kerja IoMT, data yang dikumpulkan dari perangkat medis yang terhubung diproses dan dianalisis untuk meningkatkan kemampuan pengambilan keputusan klinis[3]

Namun, seiring meningkatnya konektivitas, risiko serangan siber terhadap infrastruktur IoMT juga semakin tinggi. Salah satu sumber data yang digunakan untuk menganalisis dan mendeteksi ancaman tersebut adalah dataset CICIoMT2024. Dataset ini dirancang khusus untuk kebutuhan analisis keamanan siber di lingkungan IoMT dan memuat berbagai jenis serangan seperti DDoS, malware, serta informasi karakteristik lalu lintas jaringan yang berguna dalam pengembangan sistem deteksi serangan yang lebih andal[4], [5]. Metode pembelajaran mesin (machine learning) adalah salah satu solusi yang dapat digunakan untuk mengatasi masalah keamanan IoMT[6].

Metode pemilihan fitur Chi-Square dan penerapan algoritma pengklasifikasian seperti Random Forest yang dapat digunakan untuk mengidentifikasi fitur-fitur yang relevan untuk deteksi serangan[7], sehingga model dapat bekerja lebih efisien dengan memfokuskan perhatian pada fitur-fitur tertentu. Penulis akan menguji dan membandingkan efektivitas kedua algoritma untuk mendeteksi serangan dengan menggunakan dataset CICIoMT2024, yang dirancang khusus untuk analisis keamanan dalam lingkungan IoMT[8]. Tujuan penelitian ini adalah untuk mengidentifikasi fitur-fitur mana saja yang paling relevan dalam mendeteksi serangan dalam hal ini menggunakan metode Chi- Square serta mengevaluasi seberapa efektif algoritma Random Forest dalam mendeteksi serangan pada jaringan Internet of Medical Things (IoMT)[9]. Penelitian terkait yang pernah di lakukan oleh Sajjad et al, pada pengembangan dataset multi-protokol untuk menilai keamanan perangkat IoMT yang diberi nama CICIoMT2024, yang mana untuk mencapai hal ini, 18 serangan berbeda dilakukan terhadap topologi IoMT yang terdiri dari 40 perangkat IoMT.

Selain itu, tiga protokol ditargetkan dengan mempertimbangkan karakteristik operasi perawatan kesehatan (yaitu Wi-Fi, MQTT, dan Bluetooth). Tujuan utamanya adalah untuk berkontribusi pada state-of-the-art yang ada dengan mendefinisikan garis dasar pelengkap yang mendukung para peneliti dalam menyelidiki dan mengembangkan solusi baru

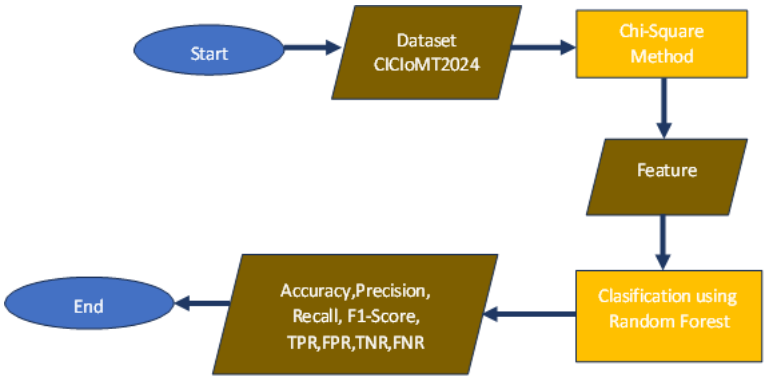
Journal of Informatics Management and Information Technology
ISSN 2774-4744 (Media Online)
Vol 5, No 2, April 2025| Hal 95-103
DOI: 10.47065/jimat.v5i2.478
https://hostjournals.com/jimat

untuk keamanan siber dalam layanan kesehatan dan operasi IoMT. Hasil yang diperoleh berupa dataset multi-protokol baru yang berisi lebih dari 230.000 rekaman dalam format PCAP dan CSV. Teknik pembelajaran mesin diterapkan pada dataset ini, menghasilkan akurasi hingga 99,55%[10]. Penelitian berikutnya yang dilakukan oleh Julian Rafapa dan A. Konokix pada tahun 2024 mengenai pengumpulan dan ekstraksi fitur dari dataset ransomware, pemodelan dengan teknik Aggregated Random Forest, dan evaluasi kinerja model menggunakan berbagai metrik. Hasil menunjukkan bahwa model tersebut mencapai akurasi hingga 94,2% lebih tinggi daripada metode lain seperti SVM dan neural network—dan juga menggunakan sumber daya sistem dengan lebih efisien[11].

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Metode Analisis yang digunakan dalam penelitian ini adalah metode analisis anomali. Metode analisis anomali adalah salah satu metode yang paling umum digunakan untuk mendeteksi serangan jaringan[12]. Metode ini menemukan perilaku atau aktivitas yang tidak biasa dalam lalu lintas jaringan. Perilaku ini dapat berupa peningkatan permintaan koneksi dari sumber yang tidak dikenal atau peningkatan lalu lintas dari sumber yang tidak dikenal. Tahapan penelitian yang dilakukan dapat dilihat pada Gambar 1, dibawah ini :



Gambar 1. Tahapan Proses Penelitian

Dalam penelitian ini, tahapan-tahapan dari analisis kebutuhan antara lain:

a. Mulai (Start).
b. Dataset CICIoMT2024.
   Pada tahap ini, penulis mengumpulkan data yang diperlukan yaitu dataset CICIoMT2024, yang bertujuan untuk mendapatkan informasi tentang kebutuhan dari penulis dan analisis yang dibutuhkan [13].
c. Chi-Square method.
   Langkah selanjutnya adalah memilih fitur yang relevan, dalam analisis ini fitur yang dipilih menggunakan metode Chi-Square. Chi-Square adalah teknik statistik yang digunakan untuk menentukan apakah ada hubungan signifikan antara algoritma random forest [14].
d. Feature.
   Setelah menggunakan fitur Chi-Square kemudian fitur tersebut digunakan untuk melatih model Machine Learning menggunakan algoritma Random Forest[15].
e. Classification Using Random forest.
   Selanjutnya pada tahap klasifikasi, menggunakan algoritma random forest yang digunakan untuk membangun model prediktif. Random Forest: Merupakan metode dari algoritma machine learning yang membangun model secara bertahap dengan menambah pohon keputuhan baru yang mengoreksi kesalahan dari model sebelumnya[16]. Algoritma ini akan dilatih menggunakan dataset CICIoMT2024 yang telah diproses dan fitur yang telah dipilih.
f. Accuracy, FPR, Recall and Sensitivity.
   Selanjutnya untuk mengevaluasi performa model, digunakan beberapa metrik seperti akurasi, FPR, Recall, dan sensitivitas. Selain itu, K-Fold Cross Validation juga digunakan untuk menghindari overfitting dan memastikan model bekerja stabil pada data yang berbeda. Dalam penelitian ini, digunakan nilai K = 10, yang berarti data dibagi menjadi 10 bagian, dan pelatihan serta pengujian dilakukan sebanyak 10 kali, masing-masing dengan satu bagian sebagai data uji dan sisanya sebagai data latih. Pemilihan nilai K = 10 didasarkan pada praktik umum dalam validasi model yang memberikan keseimbangan antara bias dan varians.
g. End
   Pada tahapan hasil pengukuran dapat menunjukkan performa dari model yang dibuat.

2.2 Metode Pengujian

Metode pengujian dalam penelitian ini bertujuan untuk menilai kinerja sistem dalam mendeteksi serangan pada Internet of Medical Things (IoMT) dengan menerapkan pemilihan fitur Chi-Square serta algoritma Random Forest[17]. Pengujian

Journal of Informatics Management and Information Technology
ISSN 2774-4744 (Media Online)
Vol 5, No 2, April 2025| Hal 95-103
DOI: 10.47065/jimat.v5i2.478
https://hostjournals.com/jimat

pertama dilakukan pada dataset untuk memastikan bahwa data mencakup berbagai jenis serangan dan data normal yang representative[18]. Evaluasi dataset dilakukan berdasarkan kualitas, distribusi kelas, dan kelengkapan fitur setelah melalui tahap preprocessing dan pemilihan fitur menggunakan Chi-Square[19].

Evaluasi model klasifikasi dilakukan dengan melatih algoritma Random Forest menggunakan data yang fiturnya sudah dipilih[20]. Model ini diuji melalui cross-validation untuk memastikan hasil yang konsisten, kemudian diuji pada data testing yang terpisah. Performa model diukur dengan metrik seperti[8]:

a. Akurasi : $\frac{TP+TN}{TP+TN+FP+FN}$ (1)

b. Precision : $\frac{TP}{TP+FP}$ (2)

c. F1-Score : $2\left(\frac{Precission*Recall}{Precission+Recall}\right)$ (3)

d. False Positive Rate (FPR) : $\frac{FP}{FP+TN}$ (4)

e. True Positive Rate (TPR) : $\frac{TP}{TP+FN}$ (5)

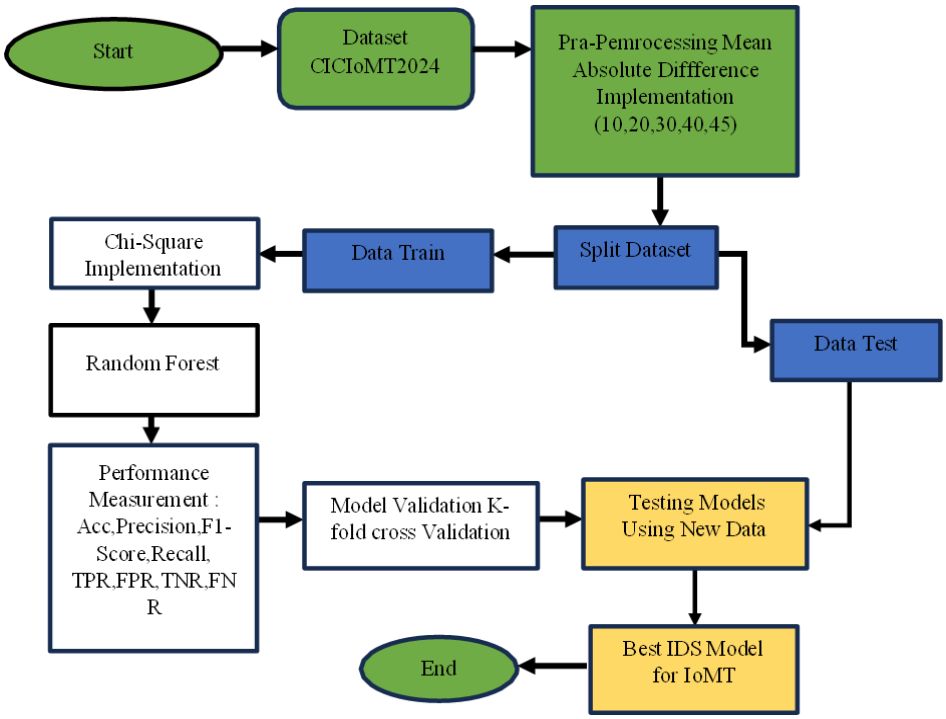
f. True Negative Rate (TNR) : $\frac{TN}{TP+FN}$ (6)

g. False Negative Rate (FNR) : $\frac{FN}{TP+FN}$ (7)

2.3 Metode Pengolahan Data

Sebelum dibagi menjadi data latih dan data uji, data yang dikumpulkan diolah melalui proses pembersihan, encoding, dan normalisasi[21]. Perhitungan Chi-Square digunakan untuk memilih fitur yang signifikan untuk digunakan dalam pelatihan model Random Forest menggabungkan berbagai pohon keputusan. Untuk mengevaluasi kinerja kedua model, metrik seperti akurasi, ketepatan, dan recall, digunakan[22].

Hasilnya menunjukkan bahwa, dengan membandingkan keunggulan masing-masing algoritma, pemilihan fitur yang tepat dapat meningkatkan kinerja deteksi serangan. Pada Gambar 2 merupakan tahapan-tahapan dalam pengolahan untuk menentukan model terbaik:

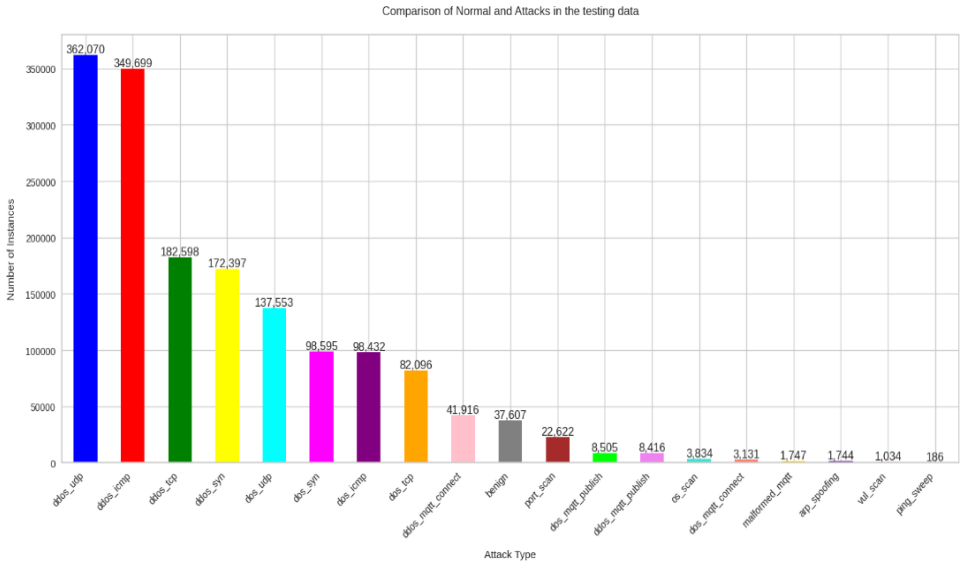


Gambar 2. Flowchart Menentukan Model Terbaik

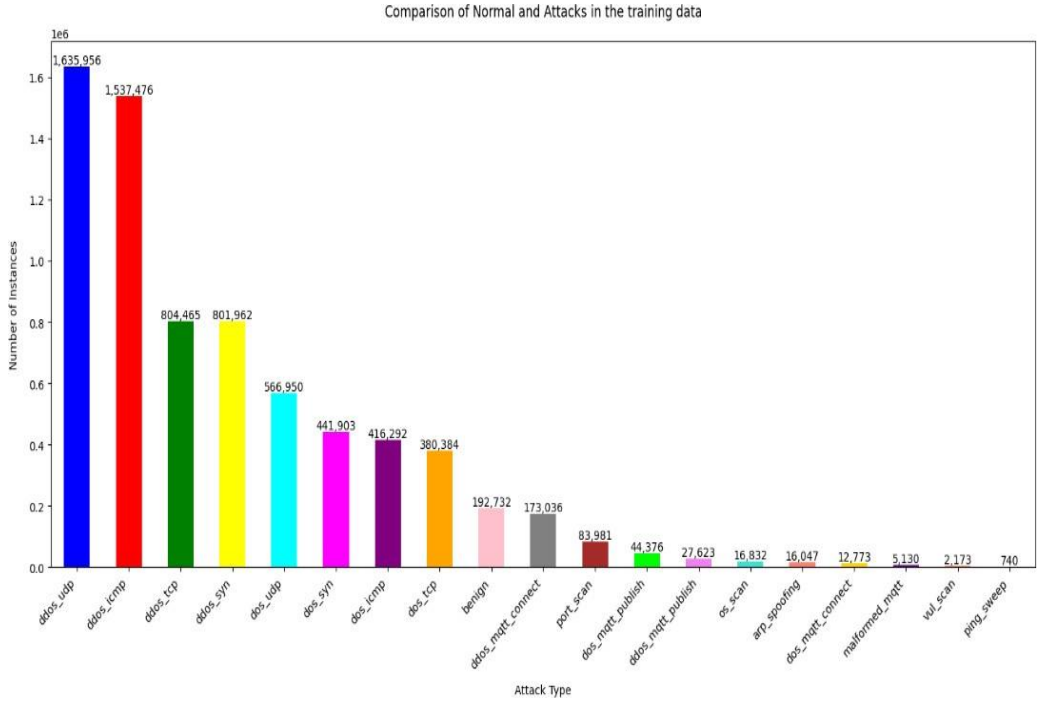
Dalam penelitian ini, tahapan-tahapan dari pengolahan data antara lain :

a. Start
b. Dataset CICIoMT2024
   Pada tahap ini, penulis mengumpulkan data yang diperlukan yaitu dataset CICIoMT2024, yang bertujuan untuk mendapatkan informasi tentang kebutuhan dari penulis dan analisis yang dibutuhkan, lalu menentukan class attack dari testing dan training data dalam dataset seperti Gambar 3 dan Gambar 4 dibawah ini

Journal of Informatics Management and Information Technology
ISSN 2774-4744 (Media Online)
Vol 5, No 2, April 2025| Hal 95-103
DOI: 10.47065/jimat.v5i2.478
https://hostjournals.com/jimat



Gambar 3. Class attack testing data



Gambar 4. Class Attack Training Data

c. Pra-Pemrosesan

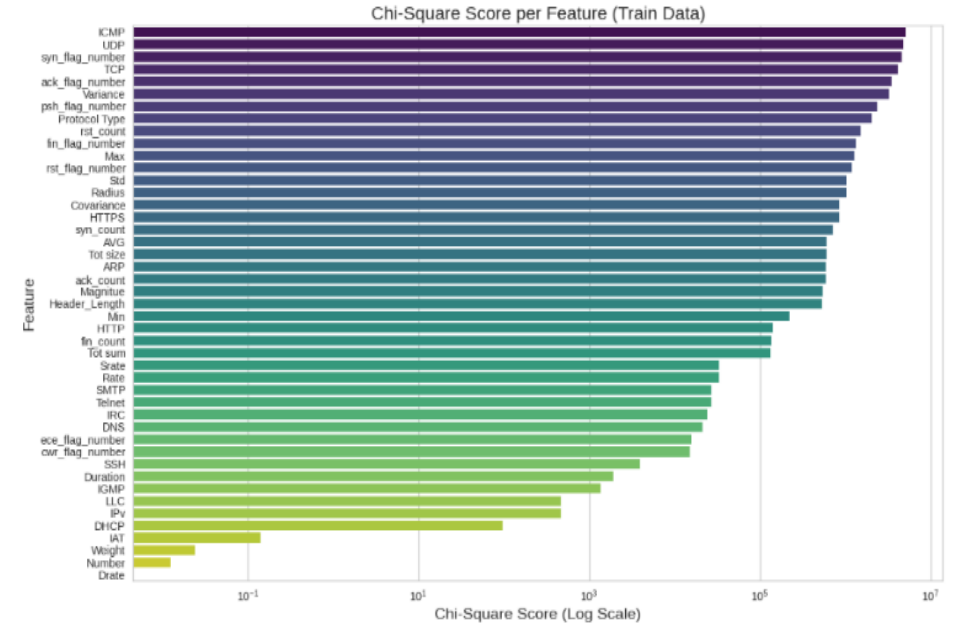
Setelah dataset dikumpulkan selanjutnya adalah pra-pemrosesan dengan menentukan nilai dari setiap fitur berdasarkan chi-square yang merupakan metode yang digunakan dalam analisis dataset CICIoMT2024[23, seperti data train pada Tabel 1 dan Gambar 5 dibawah ini .

Tabel 1. Bobot Nilai Fitur Pada Data Train

Rank	Feature	Chi-Square Score
1	ICMP	5.148480e+06
2	UDP	4.807625e+06
3	syn_flag_number	4.626391e+06
4	TCP	4.082403e+06
5	ack_flag_number	3.439229e+06
…	…	…
…	…	…
40	IPv	4.638516e+02
41	DHCP	9.663994e+01
42	IAT	1.419846e-01

Journal of Informatics Management and Information Technology
ISSN 2774-4744 (Media Online)
Vol 5, No 2, April 2025| Hal 95-103
DOI: 10.47065/jimat.v5i2.478
https://hostjournals.com/jimat

43	Weight	2.407115e-02
44	Number	1.234942e-02
45	Drate	0.000000e+00

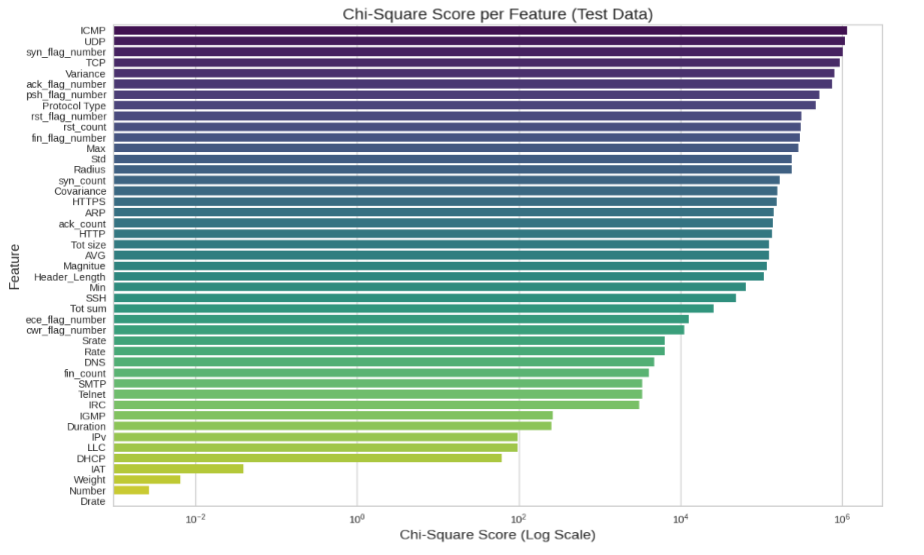


Gambar 5. Grafik Train Data

Selanjutnya untuk bobot nilai dari data test bisa dilihat pada Tabel 2 Gambar 6 dibawah ini :

Tabel 2. Bobot nilai fitur pada data test

Rank	Feature	Chi-Square Score
1	ICMP	1.162772e+06
2	UDP	1.083855e+06
3	syn_flag_number	1.007855e+06
4	TCP	9.330342e+05
5	Variance	8.084144e+05
…	…	…
…	…	…
40	LLC	9.691186e+01
41	DHCP	6.141445e+01
42	IAT	3.971906e-02
43	Weight	6.525852e-03



Gambar 6. Grafik *Test* data

ISSN 2774-4744 (Media Online)
Vol 5, No 2, April 2025| Hal 95-103
DOI: 10.47065/jimat.v5i2.478
https://hostjournals.com/jimat

d. Split dataset
Selanjutnya untuk mengembangkan model Machine Learning, split dataset adalah proses membagi dataset menjadi beberapa subset (data latih dan data uji) untuk memastikan bahwa model dapat dilatih, dan diuji secara efisien.
e. Data Train
Selanjutnya setelah data dibagi menjadi data test dan train sebelum melakukan data test, lakukan dulu data train, untuk melatih model pembelajaran mesin, ini adalah subset dataset yang membantu algoritma mengenali pola dan hubungan di dalam data.
f. Chi-Square Implementation
Langkah selanjutnya adalah pemilihan fitur. Ini akan membantu mengurangi ukuran kumpulan data dan meningkatkan kinerja model pembelajaran mesin. Fitur ini akan menentukan bobot dari feature Chi- Square mana yang lebih relevan[24].
g. Random Forest
Selanjutnya menggunakan algoritma untuk menentukan hasil seberapa relevan algortima tersebut.
h. Performance measurement : Acc, Precision, Recall, F1-Score, TPR, FPR,TNR,FNR[25].
Selanjutnya mengukur model dengan performa matrik seperti acc, precision, F1-Score, Recall, TPR, FPR, TNR, FNR.
i. Model Validation K-Fold Cross Validation
Selanjutnya model diuji dengan K-Fold Cross Validation untuk menentukan hasil yang konsisten .
j. Testing Models Using new data
Setelah diuji dengan model K-Fold Cross Validation data testing menjadi terpisah dan menghasilkan data baru yang kemudian kita uji data testing terbaru.
k. Best IDS Model for IoMT
Setelah kita melakukan tahapan-tahapan diatas, kita akan mendapatkan model IDS terbaik untuk Internet of Medical Things (IoMT) dalam menjaga keamanan data pasien dan perangkat medis.
l. End

3. HASIL DAN PEMBAHASAN

3.1 Hasil

Penelitian ini menggunakan metode chi-square untuk pemilihan fitur yang relevan, kemudian melatih fitur tersebut berdasarkan peringkat dengan bobot terbaik hasil dari chi-square. Sebelum melakukan pemilihan fitur penulis melatih model machine learning menggunakan algoritma Random Forest tanpa melakukan pemilihan fitur. Tabel 3 berikut menunjukkan hasil peforma model machine learning tanpa metode chi-square.

Tabel 3 Akurasi Model Machine Learning Tanpa Pemilihan Fitur Untuk Data Latih

Classifier	Accuracy
Random Forest	100%

Pengukuran nilai precision, F-1-Score, TPR,FPR,FNR, dan TNR model machine learning tanpa pemilihan fitur dapat dilihat pada Tabel 4 untuk algoritma Random Forest.

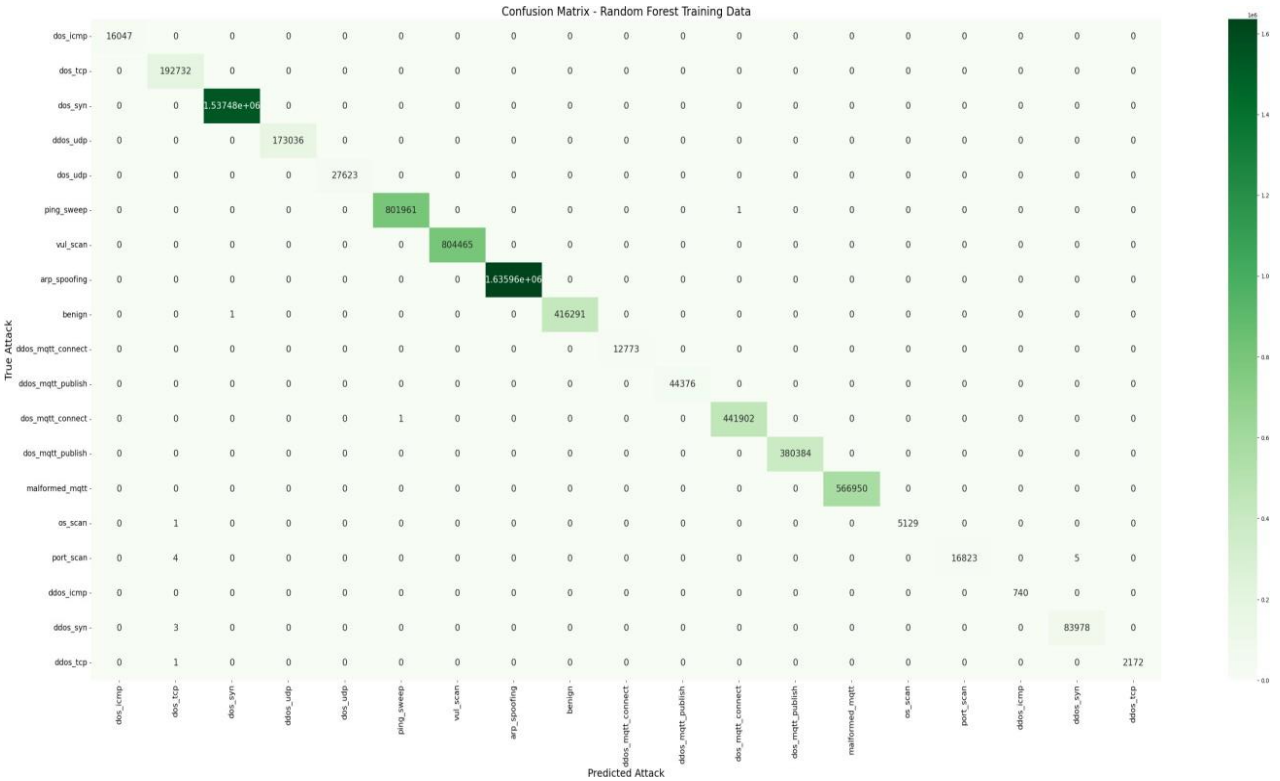
Tabel 4. Peforma Model Random Forest Tanpa Pemilihan Fitur Untuk Data Latih

Class	Precision	Recall	F1- Score	FPR	TPR	FNR	TNR
arp_spoofing	1.00	1.00	1.00	0.0000	1.0000	0.0000	1.0000
Benign	1.00	1.00	1.00	0.0000	1.0000	0.0000	1.0000
ddos_icmp	1.00	1.00	1.00	0.0000	1.0000	0.0000	1.0000
ddos_mqtt_connect	1.00	1.00	1.00	0.0000	1.0000	0.0000	1.0000
ddos_mqtt_publish	1.00	1.00	1.00	0.0000	1.0000	0.0000	1.0000
ddos_syn	1.00	1.00	1.00	0.0000	1.0000	0.0000	1.0000
ddos_tcp	1.00	1.00	1.00	0.0000	1.0000	0.0000	1.0000
ddos_udp	1.00	1.00	1.00	0.0000	1.0000	0.0000	1.0000
dos_icmp	1.00	1.00	1.00	0.0000	1.0000	0.0000	1.0000
dos_mqtt_connect	1.00	1.00	1.00	0.0000	1.0000	0.0000	1.0000
dos_mqtt_publish	1.00	1.00	1.00	0.0000	1.0000	0.0000	1.0000
dos_syn	1.00	1.00	1.00	0.0000	1.0000	0.0000	1.0000
dos_tcp	1.00	1.00	1.00	0.0000	1.0000	0.0000	1.0000
dos_udp	1.00	1.00	1.00	0.0000	1.0000	0.0000	1.0000
malformed_mqtt	1.00	0.9998	1.00	0.0000	0.9998	0.0002	1.0000
os_scan	1.00	0.9995	1.00	0.0000	0.9995	0.0005	1.0000
ping_sweep	1.00	1.00	1.00	0.0000	1.0000	0.0000	1.0000

Journal of Informatics Management and Information Technology
ISSN 2774-4744 (Media Online)
Vol 5, No 2, April 2025| Hal 95-103
DOI: 10.47065/jimat.v5i2.478
https://hostjournals.com/jimat

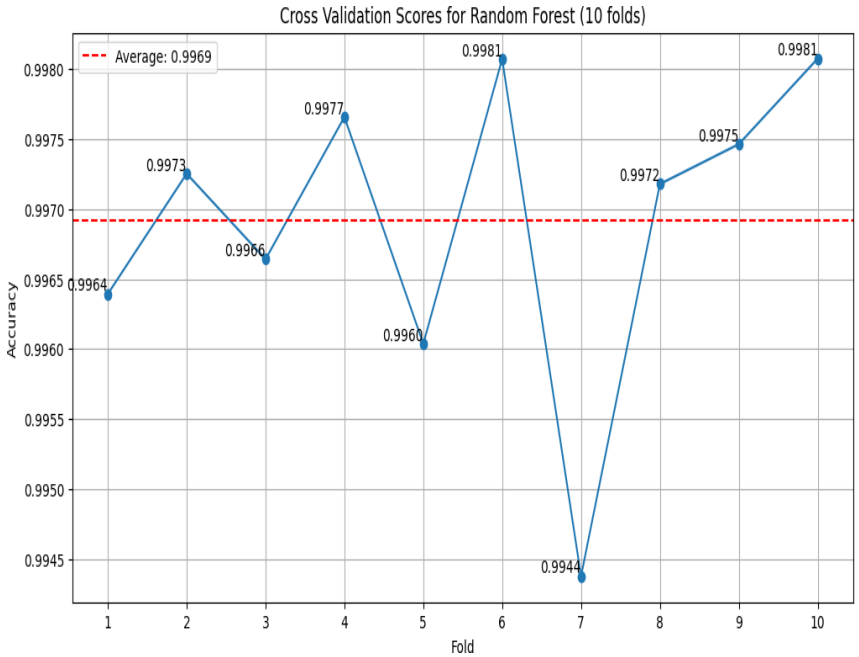
port_scan	1.00	1.00	1.00	0.0000	1.0000	0.0000	1.0000
vul_scan	1.00	0.9995	1.00	0.0000	0.9995	0.0005	1.0000

Gambar hasil confusion matrix dari algoritma random forest yang dilatih menggunakan dataset tanpa pemilihan fitur dapat dilihat pada Gambar 7, berikut ini:



Gambar 7. Confusion Matrix Untuk Algoritma Random Forest Tanpa Pemilihan Fitur Pada Data Latih

Dalam penelitian ini penulis menggunakan cross validation untuk memvalidasi model yang dibuat. Gambar 8 berikut menunjukkan hasil cross validasi dengan 10 fold untuk algoritma random forest menggunakan dataset tanpa pemilihan fitur:



Gambar 8. Cross Validasi Untuk Algoritma Random Forest Tanpa Pemilihan Fitur

Secara akurat dalam penelitian ini, model dilatih menggunakan algoritma Random Forest (RF) dengan berbagai skenario pemilihan fitur, yaitu tanpa pemilihan fitur, menggunakan 10 fitur terbaik, 20 fitur terbaik, 30 fitur terbaik, dan 40 fitur terbaik. Hasil pelatihan menunjukkan bahawa akurasi model dengan semua fitur pada data latih mencapai 100% untuk

ISSN 2774-4744 (Media Online)
Vol 5, No 2, April 2025| Hal 95-103
DOI: 10.47065/jimat.v5i2.478
https://hostjournals.com/jimat

algoritma Random Forest. Meskipun terlihat ideal, akurasi sempurna ini justru mengindikasikan potensi overfitting, yaitu saat model terlalu menyesuaikan diri dengan data latih hingga kehilangan kemampuan generalisasi. Model kemungkinan menangkap pola-pola tidak relevan yang menyebabkan penurunan performa pada data uji.

Hasil ini tidak serta-merta menunjukkan bahwa model dapat bekerja optimal pada data uji. Pengukuran nilai Precision, F1-Score, TPR, FPR, FNR, dan TNR pada tabel-tabel terkait menunjukkan bahwa beberapa kelas memiliki nilai performa yang sangat tinggi, terutama untuk kelas benign dan ddos_mqtt_connect. Namun, terdapat kelas seperti malformed_mqtt dan os_scan yang memiliki nilai TPR lebih rendah, khususnya pada algoritma Random Forest. Kinerja tinggi pada data latih tanpa pemilihan fitur ini mengindikasikan potensi overfitting. Hal ini disebabkan oleh model yang memanfaatkan semua informasi, termasuk fitur yang mungkin kurang relevan, sehingga performa pada data uji menurun. Dengan 10 fitur terbaik, akurasi model pada data latih menurun menjadi 78% untuk RF. Penurunan ini disebabkan oleh pengurangan informasi yang tersedia bagi model. Meski demikian, fitur yang dipilih diharapkan lebih relevan dan membantu dalam mengurangi overfitting. Pada data uji, akurasi lebih rendah dengan RF mencapai 78%, menunjukkan bahwa pemilihan fitur belum optimal untuk mempertahankan akurasi tinggi. Akurasi model pada data latih meningkat menjadi 81% untuk RF, sedangkan akurasi pada data uji mencapai 78% untuk RF.

Peningkatan ini menunjukkan bahwa pemilihan 20 fitur terbaik memberikan keseimbangan antara kompleksitas model dan kinerja. Pemilihan 20 fitur memungkinkan model memanfaatkan informasi penting tanpa memasukkan fitur yang tidak relevan. Akurasi model pada data latih dengan 30 fitur terbaik mencapai 99% untuk RF, sedangkan pada data uji masing-masing mencapai 75%. Dengan 40 fitur terbaik, akurasi model pada data latih meningkat menjadi 99% untuk RF, sementara pada data uji masing-masing menjadi 75%. Hal ini menunjukkan bahwa meskipun akurasi pada data latih meningkat, performa pada data uji tidak selalu menunjukkan peningkatan signifikan, yang mengindikasikan potensi overfitting pada data latih..

3.2 Analisis Hasil Validasi

Validasi dilakukan menggunakan metode cross-validation dengan 10 fold untuk mengukur kemampuan generalisasi model. Hasil validasi menunjukkan bahwa model memiliki variasi performa antar kelas yang signifikan, tergantung pada algoritma yang digunakan dan jumlah fitur yang dipilih.

Hasil cross-validation tanpa pemilihan fitur menunjukkan bahwa model memiliki akurasi rata-rata yang sangat tinggi, namun terdapat variasi nilai TPR dan FNR antar kelas. Kelas seperti benign dan ddos_mqtt_connect menunjukkan nilai TPR mendekati 1, sementara kelas seperti malformed_mqtt dan vul_scan menunjukkan nilai TPR lebih rendah, terutama pada algoritma Random Forest.

Pada validasi dengan 10 fitur terbaik, nilai rata-rata akurasi menurun dibandingkan tanpa pemilihan fitur. Namun, nilai F1-Score dan Precision pada beberapa kelas meningkat, menunjukkan bahwa fitur-fitur terpilih lebih relevan untuk mendeteksi kelas tertentu. Penambahan jumlah tanpa seleksi fitur meningkatkan akurasi rata-rata pada data latih, namun variasi antar kelas tetap signifikan. Pada beberapa kelas seperti ddos_udp dan malformed_mqtt, nilai TPR dan FNR menunjukkan peningkatan tanpa seleksi fitur terbaik.

4. KESIMPULAN

Hasil analisis performa dan validasi menunjukkan bahwa pemilihan fitur yang tepat sangat memengaruhi kinerja model, terutama dalam mengurangi overfitting dan meningkatkan kemampuan generalisasi. Algoritma Random Forest terbukti stabil dalam berbagai skenario pemilihan fitur. Namun, menentukan jumlah fitur optimal tetap menjadi tantangan. Tanpa seleksi fitur, model menunjukkan akurasi 100% pada data latih, namun kondisi ini mengidikasikan overfitting karena performa menurun saat diuji dengan data baru. Oleh karena itu, meskipun akurasinya tinggi, model tanpa seleksi fitur tidak memberikan keseimbangan yang baik antara akurasi dan generalisasi. Sebaliknya, penggunaan 20 fitur terbaik menunjukkan keseimbangan yang lebih baik, dengan akurasi 81% pada data latih dan 78% pada data uji. Ini menunjukkan bahwa seleksi fitur membantu mengurangi overfitting dan meningkatkan kemampuan model untuk mengenali pola pada data yang belum pernah dilihat.

REFERENCES

[1] A. Jalil, A. Homaidi, and Z. Fatah, “Implementasi Algoritma Support Vector Machine Untuk Klasifikasi Status Stunting Pada Balita,” *G-Tech: Jurnal Teknologi Terapan*, vol. 8, no. 3, pp. 2070–2079, Jul. 2024, doi: 10.33379/gtech.v8i3.4811.

[2] G. Sultan Aljumaie, G. Hisham Alzeer, R. Khaild Alghamdi, H. Alsuwat, and E. Alsuwat, “Modern Study on Internet of Medical Things (IOMT) Security,” *IJCSNS International Journal of Computer Science and Network Security*, vol. 21, no. 8, p. 254, 2021, doi: 10.22937/IJCSNS.2021.21.8.34.

[3] G. Thamilarasu, A. Odesile, and A. Hoang, “An intrusion detection system for internet of medical things,” *IEEE Access*, vol. 8, pp. 181560–181576, 2020, doi: 10.1109/ACCESS.2020.3026260.

[4] S. BALARAMAN, “Comparison of Classification Models for Breast Cancer Identification using Google Colab,” May 20, 2020. doi: 10.20944/preprints202005.0328.v1.

[5] J. Nayak, S. K. Meher, A. Souri, B. Naik, and S. Vimal, “Extreme learning machine and bayesian optimization-driven intelligent framework for IoMT cyber-attack detection,” *Journal of Supercomputing*, vol. 78, no. 13, pp. 14866–14891, Sep. 2022, doi: 10.1007/s11227-022-04453-z.

ISSN 2774-4744 (Media Online)
Vol 5, No 2, April 2025| Hal 95-103
DOI: 10.47065/jimat.v5i2.478
https://hostjournals.com/jimat

[6] M. Mukhopadhyay, S. Banerjee, and C. Das Mukhopadhyay, "Internet of Medical Things and the Evolution of Healthcare 4.0: Exploring Recent Trends," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 6, no. 2, pp. 182–194, Apr. 2024, doi: 10.35882/jeeemi.v6i2.402.
[7] M. Labib Mu'tashim *et al.*, "Klasifikasi Ketepatan Lama Studi Mahasiswa Dengan Algoritma Random Forest Dan Gradient Boosting (Studi Kasus Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta)," 2023.
[8] M. Eki Riyadani, "Sistem Keamanan Untuk Otorisasi Pada Smart Home Menggunakan Pengenalan Wajah Dengan Library OpenCV.", J. Sistem Komputer dan Kecerdasan Buatan, vol 5, no 2, 2022
[9] A. F. Amiri, H. Oudira, A. Chouder, and S. Kichou, "Faults detection and diagnosis of PV systems based on machine learning approach using random forest classifier," *Energy Convers Manag*, vol. 301, Feb. 2024, doi: 10.1016/j.enconman.2024.118076.
[10] S. Dadkhah *et al.*, "CICIoMT2024: Attack Vectors in Healthcare devices-A Multi-Protocol Dataset for Assessing IoMT Device Security," 2024, doi: 10.20944/preprints202402.0898.v1.
[11] J. Rafapa and A. Konokix, "Ransomware Detection Using Aggregated Random Forest Technique with Recent Variants," 2024, doi: 10.22541/au.172426891.14153527/v1.
[12] U. Tariq, I. Ullah, M. Yousuf Uddin, and S. J. Kwon, "An Effective Self-Configurable Ransomware Prevention Technique for IoMT," *Sensors*, vol. 22, no. 21, Nov. 2022, doi: 10.3390/s22218516.
[13] S. E. Ali, N. Tariq, F. A. Khan, M. Ashraf, W. Abdul, and K. Saleem, "BFT-IoMT: A Blockchain-Based Trust Mechanism to Mitigate Sybil Attack Using Fuzzy Logic in the Internet of Medical Things," *Sensors*, vol. 23, no. 9, May 2023, doi: 10.3390/s23094265.
[14] A. Fahad Rasheed, M. Zarkoosh, S. Sabah Al-Azzawi, "The Impact of Feature Selection on Malware Classification Using Chi-Square and Machine," 2023, doi: 10.1109/ICCCE58854.2023.10246084ï.
[15] W. B. Santosa, A. Syukur, and P. Purwanto, "Pemilihan Fitur Menggunakan Algoritma Chi-Square Dan Particle Swarm Optimization (PSO) Untuk Meningkatkan Kinerja Deep Neural Network Pada Deteksi Penyakit Diabetes," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 1, p. 488, Jan. 2024, doi: 10.30865/mib.v8i1.7277.
[16] E. Hariyanti, D. Pramana Hostiadi, Y. Priyo Atmojo, I. Made Darma Susila, and I. Tangkawarow, "Analisis Perbandingan Metode Seleksi Fitur pada Model Klasifikasi Decission Tree untuk Deteksi Serangan di Jaringan Komputer" Jurnal Sistem dan Informatika (JSI), vol 18, no 2, 2024
[17] D. Upadhyay, J. Manero, M. Zaman, and S. Sampalli, "Gradient Boosting Feature Selection with Machine Learning Classifiers for Intrusion Detection on Power Grids," *IEEE Transactions on Network and Service Management*, vol. 18, no. 1, pp. 1104–1116, Mar. 2021, doi: 10.1109/TNSM.2020.3032618.
[18] R. Hireche, H. Mansouri, and A. S. K. Pathan, "Security and Privacy Management in Internet of Medical Things (IoMT): A Synthesis," *Multidisciplinary Digital Publishing Institute (MDPI)*, 2022, doi: 10.3390/jcp2030033.
[19] J. G. Lugo-Armenta and L. R. Pino-Fan, "An approach to inferential reasoning levels on the Chi-square statistic," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 20, no. 1, pp. 1–19, 2024, doi: 10.29333/EJMSTE/14119.
[20] A. Habiba, R. R. Isnanto, and J. E. Suseno, "The Effect of Chi Square Feature Selection on the Naïve Bayes Algorithm on the Analysis of Indonesian Society's Sentiment About Face-to-Face Learning During the Covid-19 Pandemic," *JST (Jurnal Sains dan Teknologi)*, vol. 12, no. 1, Mar. 2023, doi: 10.23887/jstundiksha.v12i1.51899.
[21] C. T. Nnodim, O. M. Arowolo, A. A. Ajani, S. N. Okhuegbe, B. D. Agboola, and C. O. Osueke, "Emerging Advances in the Internet of Things (IoT) Technology for Fast Response to Covid-19 Outbreak With ANOVA-K-NN Implementation." 2021
[22] A. Zalukhu *et al.*, "PERANGKAT LUNAK APLIKASI PEMBELAJARAN FLOWCHART," *Jurnal Teknologi Informasi dan Industri*, vol. 4, no. 1, 2023.
[23] S. Bani Hani and M. Ahmad, "Predicting mortality amongst Jordanian men with heart attacks using the chi-square automatic interaction detection model," *Health Informatics J*, vol. 30, no. 3, Jul. 2024, doi: 10.1177/14604582241270830.
[24] W. B. Santosa, A. Syukur, and P. Purwanto, "Pemilihan Fitur Menggunakan Algoritma Chi-Square Dan Particle Swarm Optimization (PSO) Untuk Meningkatkan Kinerja Deep Neural Network Pada Deteksi Penyakit Diabetes," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 1, p. 488, Jan. 2024, doi: 10.30865/mib.v8i1.7277.
[25] U. Tariq, I. Ullah, M. Yousuf Uddin, and S. J. Kwon, "An Effective Self-Configurable Ransomware Prevention Technique for IoMT," *Sensors*, vol. 22, no. 21, Nov. 2022, doi: 10.3390/s22218516.